DOCUMENT RESUME

ED 064 324 TM 001 489

AUTHOR Lynch, Daniel O.; Smith, Billie C.
TITLE To Change or Not to Change Item Responses When Taking Tests: Empirical Evidence for Test Takers.
PUB DATE Apr 72
NOTE 6p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *College Students; *Educational Research; *Item Analysis; Objective Tests; Response Style (Tests); *Testing; Test Results; *Test Wiseness

ABSTRACT

Results of research conducted to ascertain the effect on test grades of changing answer choices are presented. The main questions that were examined were: (1) Does the changing of responses to test items (presumably based upon item reconsideration) result in better test scores?; (2) Is the amount of changes related to the score a person receives on the test?; (3) Is the pattern of changes related to the score a person receives on the test?; and (4) Does item difficulty correlate with the probability of changing a given item? Subjects were 178 university students taking final exams. The tests were composed of true-false and multiple-choice items, and most students had more than adequate time to reconsider items. All subjects were made aware of the nature of the research. The results are presented as to the effects of response changes on test scores, relationship of amount of changes to score, relationship of pattern of changes to score, and relationship of item difficulty to response changing. The current research underscores a previous finding that students who change their responses raise their scores; however, repeated changes on the same item did not improve their scores. The results also show that item difficulty is positively correlated with the probability of changing a given item, indicating that reasoned reconsideration of items was involved rather than chance misreadings of items. It also has bearing on amount and pattern of changes. (DB)

ED 064324

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Dr. Daniel O. Lynch - University of Wisconsin-Oshkosh
Dr. Billie C. Smith

AERA Paper,
April, 1972

To Change or Not to Change Item Responses When Taking Tests: Empirical Evidence for Test Takers

Most, if not all, college students enter a given course expecting or at least wanting to earn a commendable grade. To do this in most college courses, especially at the undergraduate level, it is necessary to do well on the tests designed by the instructor. Many of these tests are of the "objective" kind, usually true-false and/or multiple choice. When the test items are designed to be discriminating (probably most are), more than one response choice appears appealing. Consequently, the student in a high stress situation is forced to choose between what seems to him at the moment to be equally appealing alternatives.

Over the years students have often asked the authors the question, "Should one reconsider and possibly change a response to a test item or should he go with his first impulse?" After considering the question we came to the realization that our reply has been based on our own beliefs and feelings which have emerged from our own test taking experiences. That is, we really did not have a rational and valid basis for making a reply. Our search through the literature has led us to believe that other persons' responses to the same question have been on the same basis.

Many authors give advice on this point. Morgan and Deese, (1957, p. 77) after telling the students to be sure to carefully look over the paper for errors state, "When you re-read your examination, you'll probably be tempted to change some of your answers. We have some sound advice on this point. If you feel strongly that an answer should be changed, change it. On the other hand if you waver between two answers, not being able to make up your mind, don't change the answer you set down originally. A lot of research on this point has shown that, when you are guessing, your first guess, based on a careful reading is likely to be your best guess. If you change your answers when you're quite unsure of yourself, the chances are that you're doing the wrong thing. Remember, your first guess is probably your best."

Dressel and Jensen (1955, p. 33) concur by saying, "Don't change any of your answers unless you find you have made an obvious error." Frederick (1938, pp. 345-346) says, "Your first thought is generally best. This is a very good rule to follow in taking an objective test...if the student has time to think, he may forget the broader aspect which the teacher meant him to take and get mixed up by details." Ehrlich (1961, p. 276) admonishes students not to panic late in the test. He says, "You take a real risk of ruining correct answers." Armstrong (1956, p. 126) says, "If there is any doubt, leave your first answer."

Huff (1961, pp. 113-115) and Honig (1967, p. 123) seem to recommend changing answers. They both tell the student to answer the easy items (those he is sure of) first and then consider the remainder or harder ones and come to a decision. That is, answer the question. The idea of dealing with the easy ones first is to build confidence before considering the others. Honig says, "recheck all answers and never leave the exam before you have to."

It seems that the advice concerning the changing of answers on an objective test has been over-whelmingly against making a change. All of the authors advocate checking back over each item but this seems to be to

TM 001 489

look for errors of omission or marking one response when another was actually intended and the like, not to re-consider and possibly come to a different conclusion. However, Huff and Honig do not appear to concur. They would change if a different conclusion seemed better.

A check of student opinions showed that their opinions agreed well with the writings mentioned. Of more than 300 students in seven of the authors' classes informally interviewed, more than three out of four indicated by show of hands that they agreed that re-reading tests and then changing their answers based on reconsideration of the items involved would tend to lower their test scores.

With both students and written sources strongly favoring not changing responses to test items, it would seem reasonable to expect some scientific grounds on which these opinions are based. However, literature dating back over 40 years indicates just the opposite. Lehman (1928), Lowe (1929), Matthews (1929), Berrein (1939), and Reille and Briggs (1952) all report that changing answers (probably based on item reconsiderations) tends to raise scores more than it lowers them.

The purpose of the present report is to up-date and expand the results of the earlier studies, particularly the Reille and Briggs study. By reporting the responses of a different population of students and by using somewhat different research procedures, the present report should extend the applicability of the older results to today's students.

The main questions that were examined were:

1. Does the changing of responses to test items (presumably based upon item reconsideration) result in better test scores?

2. Is the amount of changes related to the score a person receives on the test?

3. Is the pattern of changes related to the score a person receives on the test?

4. Does item difficulty correlate with the probability of changing a given item?

PROCEDURE

Subjects were 178 summer-session education students taking final exams in three classes at the University of Wisconsin at Oshkosh in August, 1971. The three classes were Child Growth and Development (N=22), Basic Educational Psychology (N=100), and Educational Measurement and Evaluation (N=56). The tests were composed of true-false and multiple-choice items. The tests had 61, 64, and 75 items for the Child, Basic, and Measurement classes respectively. It can be concluded that most students had more than adequate time to reconsider items because they handed their tests in before they were called for.

Reille and Briggs (1952) report a study very similar to the present one. The prime difference between their study and the present one is that all subjects in the present study were made clearly aware of the nature of the research that would be done whereas Reille and Briggs' subjects took a final exam unaware that their response-changing behavior would be studied.

Each subject in the present study was given his test, answer sheet, and a sheet entitled "Changed Response Record Sheet." On this sheet there were vertically arranged numbers corresponding to the numbers of the test items and three blanks beside each number on which to record changes of responses. In the event that a subject changed a response, he recorded the change on this "Changed Response Record Sheet" by placing his original response in the first blank by the item and his second response in the second blank by the item. The heading for the first column of blanks was "From"; for the second column was "To"; and for the third column was "To". This indicated what a response was changed "from" and "to" in the event of one response change for an item and "to" a second time in the event of a second change for the item. The following directions were then given:

I am doing research to find the effect on test grades of changing answer choices. If you do not change a response your data sheet will be blank except for your name and the time at which you hand in your answer sheet. If you change responses on your answer sheet, record your changes on the data sheet. An example of this would be John Smith who the first time he went through the test put down an "E" as the answer to item 33. When he looked at 33 again, he decided "E" was wrong so he changed his answer to "B". He looked at 33 a third time and decided "C" was the correct answer. When he handed in his paper he had "C" as his answer on his answer sheet and E, B, C in the three blanks respectively on his data sheet. Are there any questions?

It should be noted that the above instructions are somewhat biased to favor reconsideration of the items.

RESULTS

Effects of response changes on test scores. There were a total of 294 response changes obtained from the 178 students. Fifty of the students did not change any responses. The maximum number of changed responses by one student was eight. Of the total of 294 changes, 21 responses obtained from 16 students were changed two or more times. Of the 21, three were triply changed responses. On the first change, nine of the 21 were from wrong to right, nine were from right to wrong, and three were from wrong to wrong. Only eight of the twenty-one were right on the last response. The three triply changed items were all wrong on the final response. Because it would have little effect on the results, it was decided to include the 21 first changes (nine wrong-to-right, nine right-to-wrong and three wrong-to-wrong) in the overall analysis of results.

Of the 294 response changes, 166 were from wrong-to-right, 79 were from right-to-wrong, and 49 were from wrong-to-wrong. If we ignore the 49 wrong-to-wrong changes because they do not affect a student's final score, on any of the remaining 235 responses odds were .68 that the student was improving his test score and .32 that he was lowering it. If .68 is a representative random sample proportion, then odds are about 99 in 100 that the true proportion of wrong-to-right changes is somewhere between .61 and .75 when these response changes are the result of the ordinary deliberation used in taking tests. The .68 proportion agrees well with the Reille and Briggs (1952) data reported in Table I in their study. They found 476 wrong-to-right and 224 right-to-wrong changes which gives accurate to two decimal places in the same proportion as found in the present study and which the authors of the present study presume is just an interesting coincidence.

Relationship of amount of changes to score. Because the different tests were not directly comparable, only the data from the 100 students in the Basic Educational Psychology class was used to find whether or not the amount of changes was related to the score a person received on the test. The observed correlation ($r = -.15$) indicated a non-significant tendency for people with higher scores to have less changed responses. Since it is possible that a sizeable number of both high and low scoring students did not think it relevant to reexamine test items, the data of the 28 students with zero response changes was discarded and the data on the remaining 72 was analyzed. In this case, the numbers of response changes were found to be significantly related with test scores ($r = -.27$, $P < .025$).

Relationship of pattern of changes to score. The response tendencies of the top and bottom 27% of the combined classes were compared to see if the pattern of changes for the high scorers was significantly different from that of low scorers. There was a nonsignificant tendency (x^2 (df=2) = 3.39; $p < .20$) for low scorers to do more poorly than high scorers when changing responses. The high scorers had proportionately somewhat more wrong-to-right changes but even the low scorers had more wrong-to-right changes than right-to-wrong.

Relationship of item difficulty to response changing. For the analysis relating to item difficulty, only the data from the 100 Basic Educational Psychology students was used. There was a low but significant correlation between item difficulty and the number of people who changed their responses to that item ($r = .25$; $p < .025$). This effect means that the more difficult the item was, the more likely people would change their responses to that item but only to a slight degree.

DISCUSSION AND CONCLUSIONS

Perusal of the leading educational and psychological testing texts reveals no mention of studies dealing with strategies of taking tests; yet, research has been available for over forty years indicating that reconsidering test items tends to raise scores. The current research underscores the validity of this finding that students who change their responses raise their scores. On the other hand, repeated changes on the same item did not help the 16 students who tried it in this study. Apparently for those items either the students were too confused about the items to make anything beyond chance improvement or decrement.

Probably the last three results analyzed should be discussed together. All are probably affected by severely attenuated data (too many zero scores in terms of number of changes and maximum number of changed responses per student only eight). In the research of Reille and Briggs (1952) the test was 130 items whereas the longest test in the current study was 64 items. The mean number of changes in the Reille and Briggs study was 7.8 whereas in the current study the mean number was 1.65. It is therefore not surprising that the results of this study with respect to the relationships of amount and pattern of changes to scores are in the same direction as earlier results but are not as strong.

The fact that item difficulty is positively correlated with the probability of changing a given item has great bearing on all results of this study. It indicates that it was in fact reasoned reconsideration of items that was involved rather than simply chance misreadings of items. It also has bearing on amount and pattern of changes. Presumably more

knowledgeable students are in the position that on an average all items are easier for them than they are for less knowledgeable students. This in turn explains both the relationship of amount of changes and the relationship of pattern of changes to test scores. People who knew more had less changes because they were more confident of their knowledge. People who knew more tended to be correct upon changing responses proportionately more often than those who knew less, again reflecting more confidence in their knowledge. However, it should be reemphasized that even the low scorers helped their scores more than they hindered them by reconsidering items.

The authors would like at this point to submit a cognitive hypothesis on the reason it is helpful to reconsider test items. Jarrett (1948) discounts "subliminal response tendencies." Presumably students taking exams have their memories jogged by other items or other reminiscences such that upon reconsideration of an item after having done other items they are more likely to be able to reason out the correct answer. It might be added that many times stems do give information that might be useful in answering other questions. What is being said then is that students do think and that optimum utilization of thinking processes during testing involves reconsideration of items. If much of reasoning involves internal responses which can be above or below threshold, then Jarrett's original hypothesis might be right.

Based on the "how-to-study" authors cited, the informal survey of 300 students, and the results of the current and previous studies, two main conclusions seem justified. First, for those students who do not go over the exam, both the reliability and validity of the exam is being lowered. Second, many students have been led astray by professors and peers telling them to stick to their first choices.

REFERENCES

Armstrong, William H. Study is hard work (second edition). New York: Harper, 1967.

Berrein, F. K. Are first impressions best on objective tests? School and Society, 1939, 50, 319-320.

Dressel, Edith P. and Jensen, Gladys C. How to study and read successfully. Teachers National Information Service, 273 Williams St., Longmeadow, Mass., 1955.

Ehrlich, Eugene H. How to study and get higher marks. New York: Thomas Crowell, 1961.

Frederick, Robert W. How to study handbook. New York: Appleton-Century, 1938.

Honig, Fred. Taking tests and scoring high; how to study for tests and pass them. New York: Arco, 1967.

Huff, Darrell. Score: the strategy of taking tests. New York: Appleton-Century, 1961.

Jarrett, R. F. "The extra-chance nature of changes in students' responses to objective test-items." Journal of General Psychology, 1948, 38, 243-250.

Lehman, Harvey C. Does it pay to change initial decisions in a true-false test? School and Society, 1928, 28, 456-458.

Lowe, Mary Louise and Crawford, C. C. First impression versus second thought in true-false tests. Journal of Educational Psychology, 1929, 20, 192-195.

Matthews, C. O. Erroneous first impressions on objective tests. Journal of Educational Psychology, 1929, 20, 280-286.

Morgan, Clifford T. and Deese, James. How to study. New York: McGraw-Hill, 1957.

Reille, Patricia J. and Briggs, Leslie J. Should students change their initial answers on objective-type tests? More evidence regarding an old problem. Journal of Educational Psychology, 1952, 43, 110-115. (Reprinted in Chase, Clinton I., and Ludlow, H. Glenn (editors). Readings in Educational and Psychological Measurement. New York: Houghton Mifflin, 1966.